bio-protocol Preprint

Not Peer Reviewed
Ask a question online

Calculating scores and classifications

MC Melissa Chiasson

Updated date: Nov 20, 2020

An abbreviated version of this protocol was published in eLIFE in Sep 2020

Multiplexed measurement of variant abundance and activity reveals VKOR topology, active site and human variant impact

DOI: 10.7554/eLife.58026

Detailed protocol

For each protein variant, we divided the number of reads for each variant in each bin by the total number of reads counted for that bin to calculate a bin-wise variant frequency ($F_{v,\ bin\ x}$). A variant's total frequency ($F_{v,\ total}$) in an experiment was calculated by summing its four bin frequencies. Variants with a total frequency higher than or equal to 10^{-4} were included in the analysis, those were lower frequencies were filtered out to reduce noise. A weighted average was calculated for each variant by multiplying the variant's bin frequency by its corresponding weight:

$$\text{Weighted average} = \frac{(F_{v,\ bin1} \text{ x } 0.25) + (F_{v,\ bin2} \text{ x } 0.5) + (F_{v,\ bin3} \text{ x } 0.75) + (F_{v,\ bin4} \text{ x } 1)}{(F_{v,\ bin1} + F_{v,\ bin2} + F_{v,\ bin3} + F_{v,\ bin4})}$$

All weighted averages thus range in value from 0.25 to 1.
To calculate the final abundance score, we min-max normalized this weighted average, using the median weighted average value of synonymous variants ($W_{synonymous}$), which was given a score of 1, and the median weighted average for non-terminal nonsense variants ($W_{nonsense}$) at positions 21 through 141. The abundance score for each variant followed this equation:

$$\text{Abundance score} = \frac{(W_{v} - W_{nonsense})}{(W_{synonymous} - W_{nonsense})}$$

The final abundance score for each variant was calculated by averaging the abundance scores derived from each replicate. Only variants that were scored in two or more replicate experiments were included in the analysis.

How to cite:(Readers should cite both the Bio-protocol preprint and the original research article where this protocol was used)

1. Chiasson, M. (2020). Calculating scores and classifications. Bio-protocol Preprint. bio-protocol.org/prep645.
2. Chiasson, M. A., Rollins, N. J., Stephany, J. J., Sitko, K. A., Matreyek, K. A., Verby, M., Sun, S., Roth, F. P., DeSloover, D., Marks, D. S., Rettie, A. E. and Fowler, D. M.(2020). Multiplexed measurement of variant abundance and activity reveals VKOR topology, active site and human variant impact. eLIFE. DOI: 10.7554/eLife.58026

Copyright: Content may be subjected to copyright.